

### **REMARKS**

We have amended the claim to more particularly point out and distinctly claims the invention. We have also added new dependent claims 23-26. Claims 1-17 and 20-26 are pending in this application.

The examiner rejected claims 1-7 and 20-22 under 35 U.S.C. §101 as directed non-statutory subject matter because the recited process (1) is not tied to a particular apparatus or machine or (2) does not transform underlying subject matter to a different state or thing. We have addressed this objection by amending claims 1 and 11 to recite utilizing a computer system to perform the recited functions.

The examiner rejected claims 1, 3-12, 17 and 20-22 under 35 U.S.C. §102(b) as anticipated by Smith et al. (Disambiguating Geographic Names in a Historical Digital Library) and Wacholder et al. (Disambiguation of Proper Names in Text) or, in the alternative, under 35 U.S.C. § 103(a) as obvious over Smith in view of Wacholder. The examiner appears to recognize that Smith does not explicitly teach “computing a value for a confidence that the selected toponym means that selected reading, wherein computing said value involves a summation over all documents in the corpus in which geo-textual correlations were identified that involved that toponym-reading pair.” So, for this aspect he turns to Wacholder to which Smith makes reference.

The examiner argues:

As taught by Wacholder, the Nominator links together variants that refer to the same entity... After the whole document collection has been processed, linked groups are merged across documents and their variants combined, e.g., if in one document “president Clinton was a variant of William Clinton, while in another document Governor Clinton was a variant of William Clinton, both are treated as variants of an aggregated William Clinton group. In this minimal sense, Nominator uses the larger context of the document collection to learn more variants for a given name... The Wacholder’s teaching reads on the claimed limitation *computing said value involves a summation over all documents in the corpus in which geo-textual correlations were identified that involved that toponym-reading pair*, e.g., by using Nominator, the computed score involves an aggregation of documents in which “Philadelphia”, “Harrisburg” and “Lancaster” were identified that involved <“Lancaster”, “Pennsylvania”>.

We disagree with the examiner's characterization of Wacholder. An aggregation of variants such as Wacholder describes is not the same as a summation over all documents, as recited in claim 1. It is not sufficient to simply look at all documents in a corpus of documents to find variants of names; one must compute a summation over those documents to satisfy the claim. Wacholder does not perform a summation of any kind. Wacholder simply identifies and links variants of a word or name so that an observation about a member of a group applies to the other linked members of that group.

Wacholder describes what she means by an aggregation and thereby makes it clear that she is not doing a summation of any kind:

Finally, Nominator groups together name variants that refer to the same entity. After information about names and their referents has been extracted from individual documents, an aggregation process combines the names collected from all the documents into a dictionary. (see page 204, bottom right column to top left column on page 205).

We also note that Wacholder is not working with toponym-reading pairs, which is the subject of claim 1. Rather, she is working with identifying canonical types, e.g. person, place, organization, etc. She states:

Finally, Nominator links together variants that refer to the same entity. Because of standard English-language naming conventions Mr. Jordan is grouped with Robert Jordan. ABA is grouped with American Bar Association as a possible abbreviation of the longer name. Each linked group is categorized by an entity type and assigned a 'canonical name' as its identifier. The canonical name is the fullest, least ambiguous label that can be used to refer to the entity.

...

After the whole document collection has been processed linked groups are merged across documents and their variants combined. (see page 205 bottom of left column to top of right column.)

In other words, if the teachings of Wacholder via the Nominator were applied to Smith, one would not have the claimed invention. Rather, Smith as modified by Wacholder would be constructing a dictionary with variants linked together and with canonical types identified for each linked group.

We have made more explicit the nature of the summation recited in claim 1 by amending claim 1 to recite a mathematical summation.

With regard to claim 10, the examiner appears to acknowledge that Smith does not explicitly teach the three numbered steps of claim 10 since he relies on Wacholder to supply these elements. We agree that Smith does not explicitly teach:

- (1) obtaining a pre-computed number for a value of a confidence that the toponym of the selected (toponym, place) pair refers to the place of the selected (toponym, place) pair, said pre-computed number derived from a statistical observation about a large corpus of documents;
- (2) determining if another toponym is present within the target document that has an associated place that is geographically related to the place referred to by the selected (toponym, place) pair; and
- (3) if a toponym is identified within the target document that has an associated place that is geographically related to the place referred to by the selected (toponym, place) pair, boosting the value of the confidence for the selected (toponym, place) pair for the target document.

We disagree, however, that Wacholder supplies these missing elements. We note, as we did above, that Wacholder describes a process for identifying the canonical form of a word. That form might be organization, person or place. The scoring mechanism that Wacholder describes is for the purpose of identifying the word's canonical form. But the canonical form of a word is not the same as a reading of a toponym. In other words, Wacholder does not deal with "a corresponding plurality of (toponym, place) pairs, wherein the place of each (toponym, place) pair of the plurality of (toponym, place) pairs identifies a geographical location or region designated by the toponym."

Wacholder makes it clear that she is concerned with the canonical form in the following passages found on page 207 left column:

Our measure of ambiguity is very pragmatic. It is based on the confidence scores yielded by heuristics that analyze a name and determine the entity types it can refer to. If the heuristic for a certain entity type (a person, for example) results in a higher confidence score (highly confident that this is a person name), we determine that the name unambiguously refers to this type. Otherwise, we choose the highest score obtained by the various heuristics.

A few simple indicators can unambiguously determine the entity type of a name, such as Mr. for a person or Inc. for an organization. More commonly, however, several

pieces of positive and negative evidence are accumulated in order to make this judgement [sic].

We have defined a set of obligatory and optional components for each entity type...

Wacholder provides no examples of determining the score for a reading of a toponym (i.e., the location designated by the toponym) and no examples of boosting that score based on discovering the presence of another toponym within the target document that refers to a geographically related place.

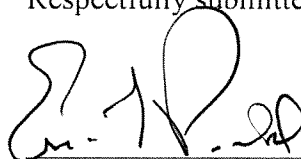
We could also find no suggestion to apply the techniques of the Nominator to score (toponym, place) pairs in the manner recited in claim 10.

For at least the reasons stated above, we believe that the claims are in condition for allowance and therefore ask the Examiner to allow them to issue.

Please apply any charges not covered, or any credits, to Deposit Account No. 08-0219, under Order No. 0113744.00124US2 from which the undersigned is authorized to draw.

Respectfully submitted,

Dated: February 19, 2010



Eric L. Prah  
Registration No.: 32,590  
Attorney for Applicant(s)

Wilmer Cutler Pickering Hale and Dorr LLP  
60 State Street  
Boston, Massachusetts 02109  
(617) 526-6000 (telephone)  
(617) 526-5000 (facsimile)